

Санкт–Петербургский государственный университет

Шарафутдинов Тимур Рустемович

**Выпускная квалификационная
работа**

***Прогнозирование фондовых индексов с использованием
искусственных нейронных сетей***

Уровень образования: бакалавриат

Направление 02.03.02 «Фундаментальная информатика и
информационные технологии»

Основная образовательная программа СВ.5003.2017 «Программирование
и информационные технологии»

Научный руководитель:

Кандидат физико-математических наук,
доцент кафедры теории систем управления
электрофизической аппаратурой
Козынченко Владимир Александрович

Рецензент:

Доктор физико-математических наук,
профессор кафедры компьютерного
моделирования и многопроцессорных
систем
Андрианов Сергей Николаевич

Санкт-Петербург

2021 г.

Оглавление

Введение	3
Постановка задачи	5
Обзор литературы	8
Глава 1. Кодировщики	11
1.1. Общая схема решения	11
1.2. GAF.....	11
1.3. Индикаторы технического анализа	12
1.4. Скользящая средняя.....	13
1.5. Моментум.....	14
1.6. Полосы Боллинджера	14
1.7. Формирование датасета.....	15
 Глава 2. Сверточные нейронные сети	 17
Глава 3. Эксперименты.....	20
3.1. Выбор индекса и источника данных	20
3.2. Программная реализация	21
3.3. Параметры эксперимента	22
3.4. Нейронная сеть	23
Выводы	25
Заключение	27
Список литературы	28
Приложения.....	31
3.1. Приложение А	31
3.2. Приложение В	35
3.3. Приложение С	

Введение

Рынок ценных бумаг имеет очень существенное влияние на экономику. Кроме того, с развитием информационных технологий, к рынку открывается доступ для простых граждан, располагающих относительно маленьким бюджетом для участия в финансовых операциях по сравнению с крупными инвесторами и корпорациями. Соответственно растет и количество участников в торговле на бирже. Для всех участников рынка операции с ценными бумагами могут стать источником для получения прибыли. Однако для успешной торговли на бирже человеку необходимо правильно определять направление движения котировок, а в идеале и значения изменений цен. Обе эти задачи не являются простыми для решения. В связи с этим возникает необходимость в совершенствовании инструментов для получения прогнозов.

На сегодняшний день широко известны два подхода к прогнозированию поведения финансовых рынков: технический и фундаментальный анализ. Суть первой группы методов заключается в выявлении закономерностей изменения цен в прошлом. Основой для поиска являются графики цен акций за определенный период. Идеей, на которой построен фундаментальный анализ, является использование доступной о компании информации, в том числе различные показатели, такие как прибыль, производство за определенный период и т.п. Однако оба подхода имеют ряд недостатков. Например, результаты прогнозов с помощью методов из области технического анализа могут дать в одной и той же ситуации несколько равновероятных прогнозов. Что касается фундаментального анализа, то одной из проблем является наличие огромного числа факторов, влияющих на изменение тренда, которые часто очень сложно выявить. Из этого, в частности, вытекает проблема зависимости методов фундаментального анализа от конкретного индекса или компании, а значит, что необходимо исследовать каждую из них по отдельности, что тоже добавляет сложности для прогнозирования.

Существует также множество исследований по построению различных математических моделей, позволяющих совершать прогнозы по ценовым рядам. Однако и эти методы имеют ряд недостатков. Например, согласно [1], проблемой методов эконометрики является то, что многие из них основаны на предположении о том, что вероятность распределения цен имеет нормальный закон распределения, что на самом деле в общем случае не является таковым.

В связи с существующими проблемами вытекает необходимость в нахождении новых методов и средств для прогнозирования поведения финансовых рынков. В частности, в последнее время проводятся исследования возможности использования методов машинного обучения для решения задачи прогнозирования. Тем не менее, несмотря на уже имеющиеся результаты исследований, показывающих возможности успешного применения методов из данной области, повышение точности прогнозирования до сих пор остается актуальной задачей.

В данной работе рассматривается один из подходов на основе сведения задачи прогнозирования временных рядов к задаче классификации с использованием GAF(Gramian angular field) и сверточных нейронных сетей. Также делаются попытки улучшения модели с помощью инструментов технического анализа, суть которых заключается в использовании индикаторов технического анализа MA(moving average), Momentum и Bollinger Bands. На их основе формируются изображения с помощью GAF и подаются на вход сверточной нейронной сети, которая решает задачу классификации.

Постановка задач

Целью выпускной квалификационной работы является разработка прикладного решения задачи прогнозирования фондовых индексов. Задача прогнозирования должна быть решена как задача трёхклассовой классификации (рост, падение, промежуточное состояние).

Для достижения поставленной цели необходимо решить следующие задачи:

1. Выбор фондового индекса и проведение разделения соответствующих ценовых рядов на тренировочный, валидационный и тестовый наборы;
2. Разработка и реализация алгоритмов предобработки данных
3. Реализация, обучение и тестирование нейронной сети для каждой модели предобработки данных
4. Проведение анализа результатов вычислительных экспериментов

Формальная постановка задачи

Пусть дан набор из K временных рядов $X = \{X_i = (x_{1i}, x_{2i}, \dots, x_{ni}), i = \overline{1, K}\}$, каждый состоит из наблюдений x_{ji} . Для каждого такого набора ставится в соответствие y_i – метка, принадлежащая одному из классов: “Рост”, “Отсутствие роста”, “Падение”.

По данному набору, который называется обучающим, требуется найти отображение:

$$f: X \rightarrow \{ "Рост", "Отсутствие роста", "Падение" \}$$

Оценка качества классификации осуществляется с помощью следующих метрик:

1. Precision(точность):

$$\frac{TP}{TP + FP}$$

2. Recall(полнота):

$$\frac{TP}{TP + FN}$$

3. F1 score(среднее гармоническое):

$$2 * \frac{precision * recall}{precision + recall}$$

4. Accuracy:

$$\frac{Correct}{All},$$

где *Correct* – количество правильно классифицированных элементов, *All* – общее число элементов в наборе.

Метрики *Precision*, *Recall* и *F1 score* считаются отдельно для каждого класса.

Пусть рассматривается класс *Q*, тогда значения *TP*, *FP*, *FN*, *TN* для данного класса можно определить следующим образом:

TP – число объектов класса *Q*, для которых система определила класс *Q*

FP – число объектов, не принадлежащих классу *Q*, для которых система определила класс *Q*

FN – число объектов класса *Q*, для которых система определила класс, отличный от *Q*

TN – число объектов, не принадлежащих классу *i*, для которых система определила класс, отличный от *Q*

С использованием значений *TP*, *FP*, *FN*, *TN* можно построить матрицу неточностей (Таблица 1).

Таблица 1 – Матрица неточностей

Класс i		Спрогнозированный класс	
		Отрицательная	Положительная
Фактический класс	Отрицательная	TN	FP
	Положительная	FN	TP

Метрика Ассигасу вычисляется для всего набора.

Обзор литературы

На данный момент существует множество книг по техническому анализу, которых подробно описываются большое количество методов анализа финансовых рынков. Например, в книге [2] приведено описание индикаторов технического анализа с последующими объяснениями их использования. В книге [3] дается описание технических индикаторов финансового рынка.

В книге [4] можно найти различные методы оценок ценных бумаг, портфельный анализ, модели активов и многое другое. А в книге [5] предлагаются методы финансового анализа, анализа финансовых отчетов.

В статье [6] продемонстрирована возможность использования метода опорных векторов(SVM) для предсказания динамики индекса NASDAQ. Были выбраны четыре признака: волатильность цен, ценовой импульс, волатильность сектора и секторный импульс. SVM решал задачу классификации, определяя, упадет или вырастет курс.

В работе [7] предложен метод, представляющий собой комбинацию ICA (Independent component analysis) и SVM для предсказания курса акций двух разных компаний, информация о которых получена из Бухарестской и Балтийской фондовых бирж. Было показано, что данная модель дает хорошую точность, если использовать методы анализа признаков.

В статье [8] представлен метод прогнозирования цен акций с использованием RNN. Предсказанные цены оказались относительно близкими к реальным, но периодически наблюдались существенные отклонения. Автор подчеркнул, что использование нейросети непростая задача и предложил варианты дальнейшей работы для увеличения точности.

В работе [9] описан метод с использованием сети прямого распространения (FNN) для предсказания курса на примере индекса NASDAQ.

В качестве архитектуры использован многослойный перцептрон. Была получена точность порядка 80%.

В статье [10] представлено описание методов перехода от одномерных данных к двумерным матрицам с помощью методов GAF (gramian angular field) и MTF (markof transition field) и возможности их использования для решения задачи классификации временных рядов.

В статье [11] показано, что с помощью сверточных нейронных сетей можно делать прогнозы с вероятностью выше 50%. Были продемонстрированы результаты использования GAF, индикаторов технического анализа MAM(moving average mapping) и DMAM(double moving average mapping), а также свечных графиков, для решения задачи классификации временных рядов.

В работе [12] продемонстрирована возможность использования сверточных нейронных сетей для классификации временных рядов. В частности, автор показал, как можно использовать комбинацию MTF и GAF.

В книге [13] содержится подробное описание методов машинного обучения, искусственных нейронных сетей, с примерами кода реализации на языке python с помощью библиотек tensorflow и scikit-learn. В книге [14] можно найти описание искусственных нейронных сетей и примеры их реализации с помощью библиотеки keras.

В работе [15] делаются попытки прогнозирования значений фондового индекса S&P500. Автор использует две модели нейронных сетей: многослойный перцептрон и рекуррентную нейронную сеть. Последняя давала наиболее точный результат.

Существуют методы прогнозирования, основанные на построении различных математических моделей. Например, в работе [16] представлены примеры успешного в условиях стабильности рынка использования

авторегрессии для предсказания курса акций крупных компаний, таких как “Сбербанк”, “Лукойл”, “Газпром” и другие. В работе [1] показано, что можно использовать модели на основе цепей Маркова для ведения успешной торговли на бирже.

Глава 1. Кодировщики.

1.1. Общая схема решения

Построение нейросетевого классификатора можно разделить на следующие этапы: формирование набора данных, предобработка данных, обучение нейросети, тестирование нейросети. Этап предобработки для построения нейросетевого классификатора решающего задачу прогнозирования фондовых индексов можно разделить на следующие шаги:

1. Формирование значений индикаторов технического анализа
2. Разбиение на фреймы
3. Расчёт меток принадлежности к классам
4. Переход к относительным ценам
5. Нормализация
6. Получение изображений методом GAF

1.2. GAF

CNN принимает на вход изображения, но фреймы – это одномерные данные. Поэтому, чтобы передать ей фрейм, необходимо преобразовать его к двумерным данным(матрице). Для этого используется метод GAF(Gramian Angular Field).

Сначала временной ряд нормализуется на отрезок $[-1; 1]$ или $[0; 1]$. Затем приводится к полярной системе координат по формуле:

$$\varphi = \arccos(x_i), -1 \leq x_i \leq 1, x_i \in \underline{X} \quad (4)$$

$$r = \frac{t_i}{N}, t_i \in N \quad (5),$$

где N – заранее заданная константа.

Далее можно построить матрицу по формуле:

$$G = \begin{bmatrix} \cos(\varphi_1 + \varphi_1) & \cdots & \cos(\varphi_1 + \varphi_n) \\ \vdots & \ddots & \vdots \\ \cos(\varphi_n + \varphi_1) & \cdots & \cos(\varphi_n + \varphi_n) \end{bmatrix} \quad (1)$$

Опираясь на формулу (1), можно получить конечную формулу для вычисления GAF (1):

$$G = \begin{bmatrix} \langle x_1, x_1 \rangle & \cdots & \langle x_1, x_n \rangle \\ \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \cdots & \langle x_n, x_n \rangle \end{bmatrix}$$

Операция $\langle \rangle$ определяется следующим образом:

$$\langle x, y \rangle = x \cdot y - \sqrt{1 - x^2} \cdot \sqrt{1 - y^2}$$

Примеры полученных изображений можно увидеть на рисунках 1 и 2.

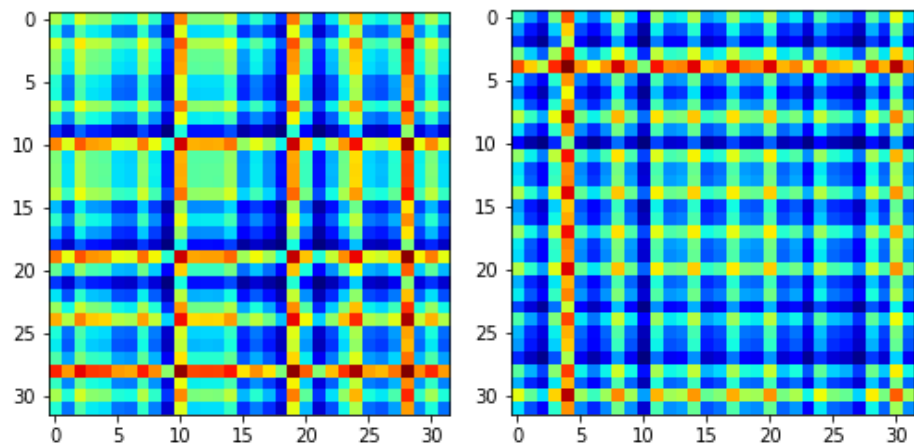


Рис. 1 и 2. Примеры изображений, полученных с помощью GAF.

1.3. Индикаторы технического анализа

Существует довольно много индикаторов технического анализа. В том числе: скользящие средние (MA – moving average), Полосы Боллинджера (Bollinger Bands), импульс (momentum), индикатор среднего направленного движения (ADX), индекс относительной силы (RSI), индикатор Стохастик

(Stochastic), индикатор Параболик (Parabolic) и другие. В данной работе будут использоваться индикаторы МА, Momentum и Полосы Боллинджера.

1.4. Скользящая средняя

Индикатор скользящей средней (МА) используется именно для определения тренда, поэтому входит в группу запаздывающих. По сути, индикатор представляет собой кривую, наклон которой показывает направление изменения цены. Для получения значения кривой в одной из точек надо просто найти среднее арифметическое за определенный промежуток времени:

$$MA = \frac{p_t + p_{t-1} + \dots + p_{t-n+1}}{n}$$

Обычно на графике вместе со скользящей средней отображается свечной график либо график цен. Пример графика изображен на рисунке 3.



Рис. 3. График цен, свечей и МА [17].

1.5. Momentum

Для вычисления значения индикатора momentum в заданной точке из значения соответствующей ей цены закрытия вычитается цена закрытия, которая была несколько дней назад. Очевидно, что могут получиться как отрицательные, так и положительные значения. При этом они колеблются в окрестности нуля. Что касается интерпретации, индикатор показывает направление тренда и сигнализирует о его возможном изменении.

$$Momentum = p_t - p_{t-n}$$

1.6. Полосы Боллинджера

Индикатор технического анализа, который состоит из двух линий, являющихся, коридором цены. Выход за пределы является сигналом для покупки или продажи. Для расчета значений полос сначала необходимо получить значения другого индикатора: скользящей средней (краткое описание и формула представлены в данной работе ранее). Затем можно получить значения полос по формулам:

$$LowerBand = MA - 2 * SD$$

$$UpperBand = MA + 2 * SD$$

$$SD = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} p_i^2}$$

Пример графика с Полосами Боллинджера представлен на рисунке 4.



Рис. 4. График цен, свечей и Bollinger Bands [17].

1.7. Формирование наборов данных

В данной работе будут использоваться 4 модели предобработки: на основе GAF и индикаторов momentum, MA, Bollinger Bands.

На первом этапе исходный ценовой ряд разбивается на три набора. Первый будет соответствовать тренировочному, второй валидационному и третий тестовому. Для разбиения необходимо указать 2 даты: первый день для тестового и первый день для валидационного наборов.

Если в качестве модели предобработки используется модель на основе одного из индикаторов, то на данном этапе получаем новые временные ряды на основе формул, представленных выше.

На следующем этапе происходит формирование фреймов длиной на 1 больше заданной. Для этого окно последовательно проходит по временному ряду с заданным шагом (передаются в качестве параметров функции) соответствующих наборов, сформированных на предыдущем шаге. Если используется модель на основе индикатора momentum, то фреймов равна заданной.

Далее формируются метки принадлежности к классу в зависимости от значения элемента ряда через некоторое время (также передается в качестве параметра функции), которое рассчитывается как:

$$x = \frac{x_l}{x_p}, \quad (2)$$

где x_l – последняя цена из фрейма, x_p – цена, относительно которой делается прогноз тренда

Метки формируются по формуле ниже:

1. 2 (увеличение цены) при $x > higherBound$
2. 1 (промежуточное состояние) при $x \geq lowerBound$ и $x \leq higherBound$
3. 0 (уменьшение цены) при $x < lowerBound$,

где x рассчитывается по формуле (2), $higherBound$ и $lowerBound$ передаются в качестве параметра функции.

Следующим этапом является получение относительных цен: отношение текущей цены к цене несколько дней назад (количество дней передается функции в качестве параметра). Если используется модель предобработки на основе momentum, то данный шаг пропускается.

Предпоследним этапом является нормализация фреймов на отрезок $[0; 1]$.

На заключительном шаге происходит преобразование фреймов к двумерным данным с использованием формул GAF.

Глава 2. Сверточные нейронные сети.

CNN имеет несколько характерных для нее типов слоев: сверточные и субдискретизирующие (пулинговые). Каждое значения выхода нейрона сверточного слоя вычисляется по формуле ([13]):

$$z_{i,j,k} = b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{k'=0}^{f_{n'}-1} x_{i',j',k'} * w_{u,v,k',k},$$

где:

- $i' = i \times s_h + u$
- $j' = j \times s_w + v$
- $z_{i,j,k}$ – выход нейрона, расположенного в строке i и столбце j в карте признаков k сверточного слоя (слоя l).
- s_h и s_w – вертикальный и горизонтальные смещения (шаги), f_h и f_w – высота и ширина рецептивного поля, а $f_{n'}$ – количество карт признаков в предыдущем слое ($l - 1$).
- $x_{i',j',k'}$ – выход нейрона, расположенного в слое $l - 1$, строка i' , столбец j' , карта признаков k' (или канал k' , если предыдущий слой является входным).
- b_k – член смещения для карты признаков k (в слое l).
- $w_{u,v,k',k}$ – вес связи между любым нейроном в карте признаков k слоя l и его входом, расположенным в строке u , столбце v и карте признаков k'

Субдискретизирующий слой агрегирует значения рецептивного поля. Принимая во внимание введенные ранее обозначения, формула для вычисления выхода нейрона субдискретизирующего слоя (average pooling):

$$z_{i,j,k} = b + \frac{1}{f_h f_w} \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} x_{i',j',k},$$

Архитектура LeNet-5(рисунок 5) представляет собой последовательность сверточных и объединяющих слоев, которая связана с несколькими полносвязными слоями. На последнем слое стоит функция активации softmax, которая принимает на вход вектор размерности 3 и возвращает вектор той же размерности, но со значениями компонент из отрезка $[0; 1]$ и суммой равной 1. Таким образом, на выходе получается набор чисел, каждое из которых интерпретируется как вероятность того, что изображение можно отнести к классу, метка которого равна порядковому номеру числа. Формула для вычисления компонент ([18]):

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^3 e^{z_k}}$$

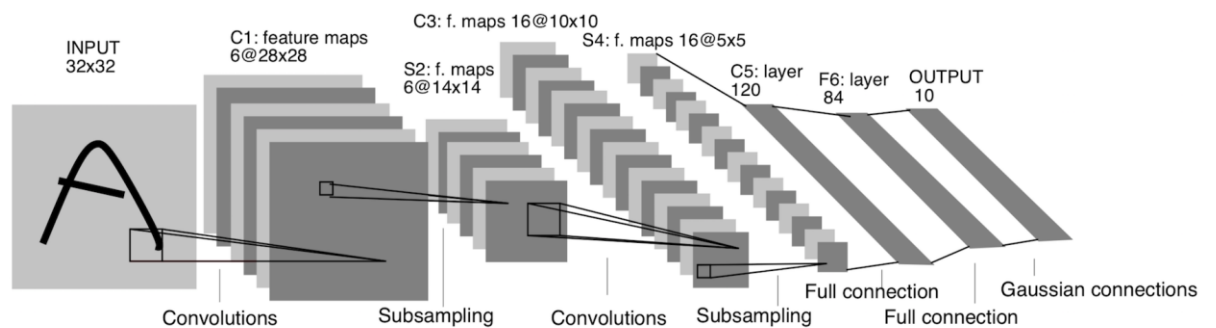


Рис. 5. Архитектура LeNet-5 [19].

Обучение нейронных сетей происходит с помощью различных градиентных методов, с использованием которых пошагово уменьшается значения функции ошибки. Расчет градиента основан на методе обратного распространения ошибки. Существует множество алгоритмов обучения нейронных сетей (оптимизаторов). В данной работе будет использоваться оптимизатор Adam, рассчитывающий на каждом шаге обучения новые значения весов по формулам ([20]):

$$g_t = \nabla_{\theta}(\theta_{t-1})$$

$$m_0 = 0$$

$$v_0 = 0$$

$$m_t = \beta_1 * m_{t-1} + (1 - \beta_1) * g_t$$

$$v_t = \beta_2 * v_{t-1} + (1 - \beta_2) * g_t$$

$$\widehat{m}_t = \frac{m_t}{(1 - \beta_1^t)}$$

$$\widehat{v}_t = \frac{v_t}{(1 - \beta_2^t)}$$

$$\theta_t = \theta_{t-1} - \alpha \frac{\widehat{m}_t}{(\sqrt{\widehat{v}_t} + \varepsilon)},$$

где θ_{t-1} – набор весов нейронной сети на шаге t , β_1 и β_2 – параметры обучения (обычно берутся 0.9 и 0.999 соответственно), α – скорость обучения, ε – константа, близкая к 0, берется чтобы предотвратить деление на 0

Глава 3. Эксперименты.

3.1. Выбор индекса и источника данных

Существенное влияние на работу нейронных сетей оказывает качество подаваемых ей на вход данных. Если они будут излишне зашумлены, то может оказаться, что нейронная сеть найдет зависимости между данными, которых на самом деле нет. Поэтому для получения наиболее достоверных результатов работы моделей желательно выбирать наиболее надежные индексы. То же самое относится и к источникам данных.

Заданным критериям удовлетворяет множество индексов и источников, в том числе индексы NASDAQ, S&P 500, Dow Jones и источник данных Yahoo Finance источник данных. Именно они и используются для обучения и тестирования сети в данной работе. Предполагается использовать два типа данных: цену закрытия и дату, соответствующую данной цене.

Графики цен закрытия для индексов приведены на рисунках 6, 7 и 8. Пунктирными линиями обозначены нижние границы валидационного и тестового множеств соответственно.

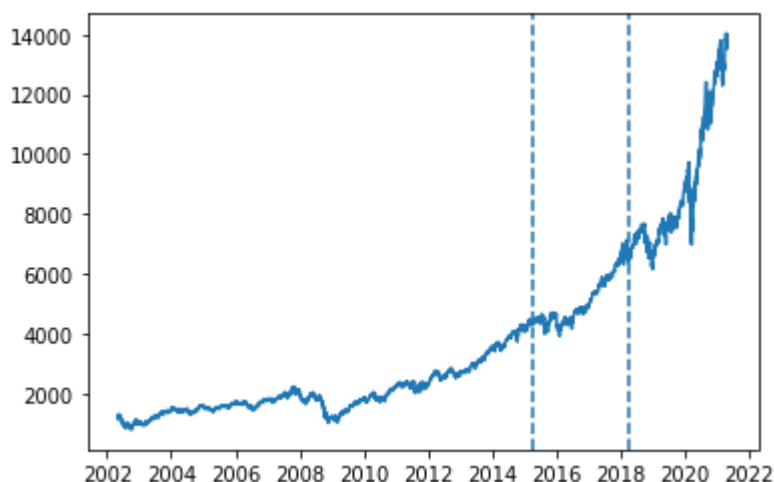


Рис. 6. График цен закрытия фондового индекса NASDAQ

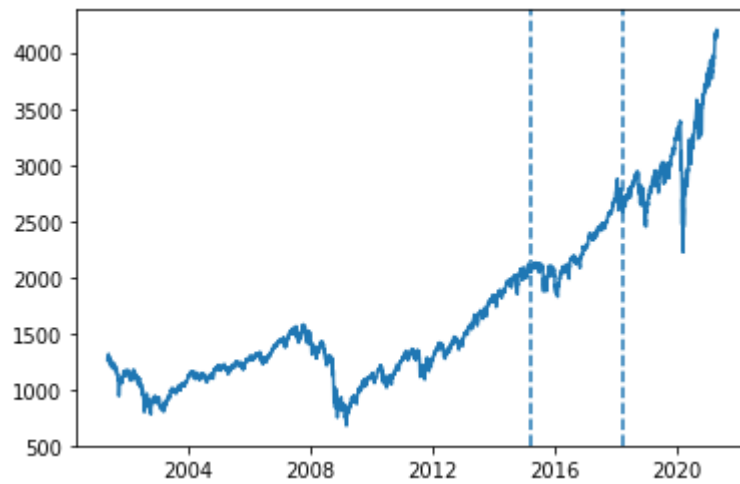


Рис. 7. График цен закрытия фондового индекса S&P500

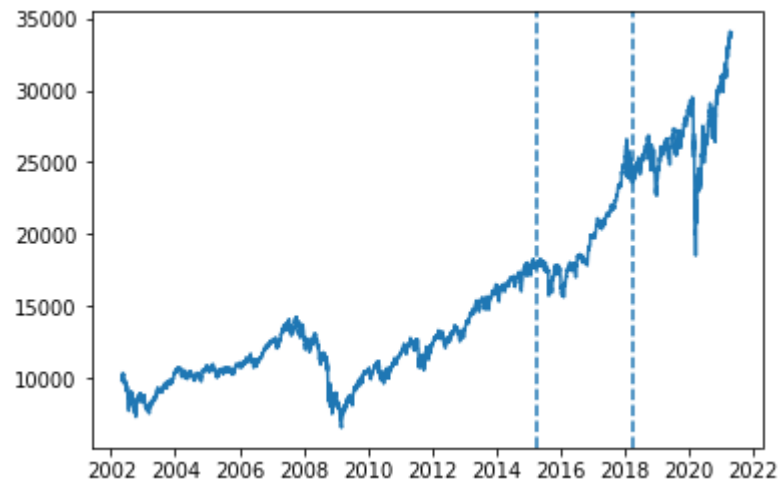


Рис. 8. График цен закрытия фондового индекса Dow Jones.

Данные охватывают период в 20 лет: с 1 апреля 1991 года по 31 марта 2021 года. Временной интервал между наблюдениями – один день.

3.2. Программная реализация

Программа была реализована на языке Python с использованием библиотек pandas, tensorflow, mplfinance и numpy.

3.3. Параметры эксперимента

Эксперименты будут проводиться на ЭВМ с процессором 2,6 GHz 2-ядерный процессор Intel Core i5 и оперативной памятью 8 ГБ 1600 MHz DDR3.

В экспериментах использовались следующие параметры:

1. Период владения активами: выражается в количестве дней вперед, на которые делается прогноз. Выбранное значение: 7.
2. Значение шага фрейма: расстояние между двумя фреймами, выраженное в количестве дней. Выбранное значение: 1.
3. Размер изображения, подаваемого на вход: для того, чтобы избежать переполнения памяти, выбирается относительно небольшой размер. Выбранное значение: 32.
4. Количество изображений в одном классе для обучающей выборки. Выбранное значение: 1000.
5. Количество изображений в одном классе для тестовой выборки. Выбранное значение: 200.
6. Число эпох для обучения: 100 для модели на основе полос Боллинджера и 50 для остальных моделей предобработки.
7. Граница для разбиения на тренировочную и валидационную: дата, относительно которой берется начало валидационной выборки. Выбранное значение: 1 апреля 2015.
8. Граница для разбиения на валидационную и тестовую выборки: дата, относительно которой берется начало тестовой выборки. Выбранное значение: 1 апреля 2018.
9. Верхняя и нижняя границы классов: значения относительно которых принимается принадлежность фрейма к данному классу. Выбранные значения зависят от индекса и подбираются так, чтобы выборки оказались сбалансированными.

10. Размер окна скользящей средней. Выбранное значение: 5.
11. Количество цен, относительно которых считается
среднеквадратическое отклонение. Выбранное значение: 5.
12. Количество дней для индикатора momentum: параметр, по
значению которого определяется, какой интервал между днями
выбирать для расчета значений индикатора. Выбранное значение: 5.

3.4. Нейронная сеть

В качестве архитектуры нейронной сети будет использоваться LeNet-5. Она имеет 3 сверточных и 2 объединяющих слоя. Архитектура реализованной нейронной сети представлена в Таблице 5.

Параметры обучения нейронной сети:

- Функция потерь – категориальная кросс-энтропия
- Оптимизатор – Adam
- Метрика – Accuracy

Таблица 2 – Архитектура нейронной сети

Слой	Размер выходного тензора	Количество обучаемых весов
conv2d_313 (Conv2D)	(32, 32, 16)	160
average_pooling2d_210	(16, 16, 16)	0
conv2d_314 (Conv2D)	(16, 16, 32)	4640
average_pooling2d_211	(8, 8, 32)	0
conv2d_315 (Conv2D)	(8, 8, 64)	18496
flatten_105 (Flatten)	(4096)	0
dense_213 (Dense)	(120)	491640
dropout_102 (Dropout)	(120)	0

dense_214 (Dense)	(3)	363
Общее число обучаемых весов: 515299		

Для определения оптимального количества эпох использовался метод раннего останова.

В обучающем наборе содержится по 1000 изображений на каждый класс. В валидационном и тестовом наборах по 200 изображений на каждый класс. Графики зависимостей и таблицы для соответствующих моделей предобработки и индексов можно посмотреть ниже.

Выводы

Из результатов вычислительного эксперимента следует, что конкретные результаты предсказания зависят от модели и от индекса, для которого делается прогноз. Итоговая таблица значений метрики Accuracy для каждого индекса и модели представлены в таблице 3.

Таблица 3 – Точность прогнозов (Accuracy) для методов GAF, MA, Momentum, Bollinger

	GAF	MA	Momentum	Bollinger bands
S&P 500	0.37	0.38	0.38	0.49
NASDAQ	0.42	0.33	0.37	0.54
DOW JONES	0.33	0.37	0.42	0.43

Подробную информацию о значениях метрик, полученных для различных индексов и методов, а также графики зависимости точности от количества эпох обучения можно найти в приложениях А, В и С.

Из Таблицы 3 видно, что нейросетевой классификатор, основанный на предобработке с использованием линий Боллинджера, можно считать успешной прогнозной моделью для индекса NASDAQ.

Для индекса S&P500 классификатор, использующий модель предобработки на основе GAF делает менее точные прогнозы по сравнению с остальными методами (0.37). MA и Momentum в данном случае не имеют существенных отличий и дают точность 0.38. Наилучшие результаты получаются при использовании метода Bollinger, точность которого 0.49.

Наиболее качественные прогнозы GAF дает для индекса NASDAQ (0.42). При этом методы MA и Momentum уступают ему. Первый показывает в данном случае наихудшие результаты (0.33). Momentum делает прогнозы

немного точнее (0.37). Особенно хорошо работает с данным индексом метод Bollinger, который делает прогнозы с точностью 0.54.

Для индекса Dow Jones самые низкие по точности прогнозы делает метод GAF (0.33). Немного лучше работает MA (0.37). Вторым по точности является метод Momentum (0.42). Наилучшим, несмотря на заметное падение в точности по сравнению с другими индексами, остается Bollinger (0.43).

Также метод Bollinger для всех используемых индексов дает наилучшие прогнозы для падения и повышения. Наименее точно он предсказывает стабильность цены.

Также можно заметить, что качество прогнозов зависит не только от модели, но и от индекса. В частности, для индексов S&P 500 и Dow Jones модели MA и Momentum показали точность лучшую, чем для NASDAQ. При этом для индекса NASDAQ эти модели делали наименее качественные прогнозы, а GAF и Bollinger, напротив, на данном индексе работали лучше всего, хотя на индексах Dow Jones и S&P 500 работали хуже.

Отдельно можно отметить метод Bollinger, который давал прогнозы, результаты которых по метрике Precision имели значения большие чем 0.5 для класса “падение” (0.6, 0.62 и 0.52 для индексов S&P 500, NASDAQ и Dow Jones соответственно). Для индекса NASDAQ метод Bollinger показал хорошую точность (0.54). Для остальных методов для каждого индекса результаты значений по метрикам Accuracy и Precision оказались неудовлетворительными.

В качестве дальнейших улучшений можно предложить другие модели предобработки на основе иных индикаторов технического анализа, а также использовать рекуррентные нейронные сети.

Заключение

В выпускной квалификационной работе:

1. Выбраны 3 фондовых индекса NASDAQ, Dow Jones и S&P 500 и проведено разделение соответствующих ценовых рядов на тренировочный, валидационный и тестовый наборы;
2. Разработаны и реализованы 3 алгоритма предобработки данных на основе метода GAF и трёх индикаторов технического анализа: MA, Momentum, Bollinger
3. Реализована, и для каждой из четырёх моделей предобработки обучена и протестирована нейронная сеть
4. Проведён анализ результатов вычислительных экспериментов по прогнозированию для трёх фондовых индексов

Список использованных источников

- [1] Е. В. Соколов Модели прогнозирования цен акций с применением функций Уолша и марковских цепей / Е. В. Соколов, Д. В. Бородин // Прикладная информатика, 2010. – 3–12 с.
- [2] Швагер Д.Д. Технический анализ. Полный курс. – Альпина Паблишер, 14-ое издание, 2018. – 804 с.
- [3] Стивен Б. А. Технический анализ от А до Я. – Омега-Л, 2010. – 376 с.
- [4] Уильям Ф. Ш., Гордон Дж. А., Джеффри В. Б. Инвестиции. Университетский учебник / Уильям Ф. Ш., Гордон Дж. А., Джеффри В. Б. – ИНФРА-М 2018. – 1028 с.
- [5] Бенджамин Г., Дэвид Д. Анализ ценных бумаг. / Бенджамин Г., Дэвид Д. – Вильям, 2017. – 880 с.
- [6] Madge S. Predicting Stock Price Direction using Support Vector Machines // Independent Work Report Spring, 2015. – 14 с.
- [7] Grogoryan H. A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA) // Database Systems Journal vol. VII, no. 1, Bucharest University of Economic Studies, Bucharest, Romania, 2016. – 12–21 с.
- [8] Лысов К. А. Прогнозирование ценовых колебаний и долгосрочных трендов на финансовых рынках магистерская диссертация. – СПбГУ, 2018. – 53 с.
- [9] Hedayati A. M. Stock market index prediction using artificial neural network / Hedayati A. M. Hedayati M. M., Morteza E. // Journal of Economics, Finance and Administrative Science, 2016. – 5 с.

[10] Zhiguang W. Encoding Time Series as Images for Visual Inspection and Classification Using Tiled Convolutional Neural Networks / Zhiguang W., Oates T., Computer Science and Electrical Engineering Department University of Maryland Baltimore County, Trajectory Based Behavior Analytics // Association for the Advancement of Artificial Intelligence, 2015. – 40–46 с.

[11] Jou-Fan C. Financial Time-series Data Analysis using Deep Convolutional Neural Network. / Jou-Fan C., Wei-Lun C., Chun-Ping H. // IEE Computer Society, 7th International Conference on Cloud Computing and Big Data, 2016. – 87–92 с.

[12] Кузнецов М. Д. Применение сверточных нейронных сетей для анализа временных рядов. – М: Московский государственный университет имени М. В., 2017. – 26 с.

[13] Жерон О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. – пер. с англ. – СПб ООО “Альфа-книга”. – 688 с.

[14] Джулли А., Пал С. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотеки Theano и TensorFlow / Джулли А., Пал С. – пер. с англ.: ДМК Пресс, 2018. – 294 с.

[15] Вакарин Д. Н. Прогнозирование цен на фондовом рынке с помощью методов машинного обучения. – ТГУ, 2019. – 58 с.

[16] Еремчук Н. А. Анализ прогнозирования ценовой динамики фондового рынка. – СПбГУ, 2016. – 114 с.

[17] Хабр [Электронный ресурс] URL: <https://habr.com/ru/post/419239/> (дата обращения: 03.04.2021)

[18] Википедия [Электронный ресурс] URL: <https://ru.wikipedia.org/wiki/Softmax> (дата обращения: 28.05.2021)

[19] Medium [Электронный ресурс] URL:
<https://congyuzhou.medium.com/lenet-5-своими-руками-b60ae3727cd3> (дата обращения: 11.04.2021)

[20] Medium [Электронный ресурс] URL:
<https://medium.com/@nishantnikhil/adam-optimizer-notes-ddac4fd7218> (дата обращения: 29.05.2021)

Приложения

Приложение А. Результаты для индекса S&P 500 GAF

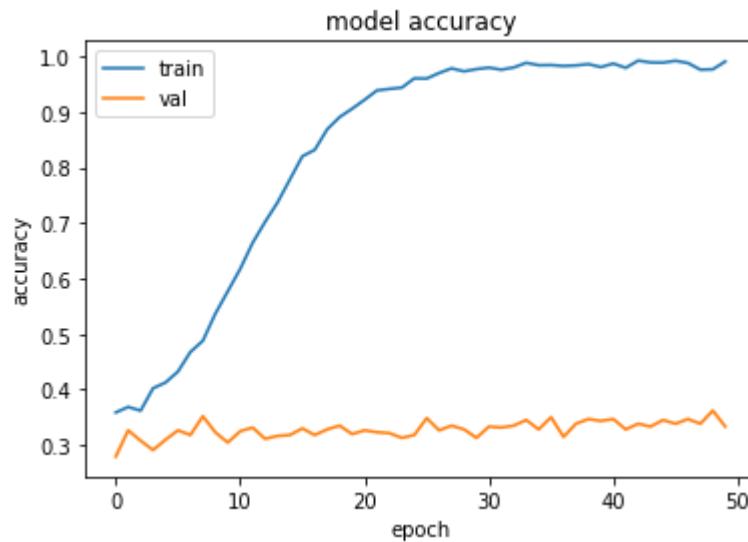


Рис. 9. График зависимости точности прогнозов GAF для индекса S&P 500 от числа эпох для обучающего и валидационного наборов.

Таблица А.1 – Значение метрик качества для метода GAF и индекса S&P 500.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.34	0.29	0.31	200
1 (Промежуточное состояние)	0.36	0.50	0.42	200
2 (Рост)	0.44	0.33	0.38	200

Полученное значение для Accuracy: 0.37.

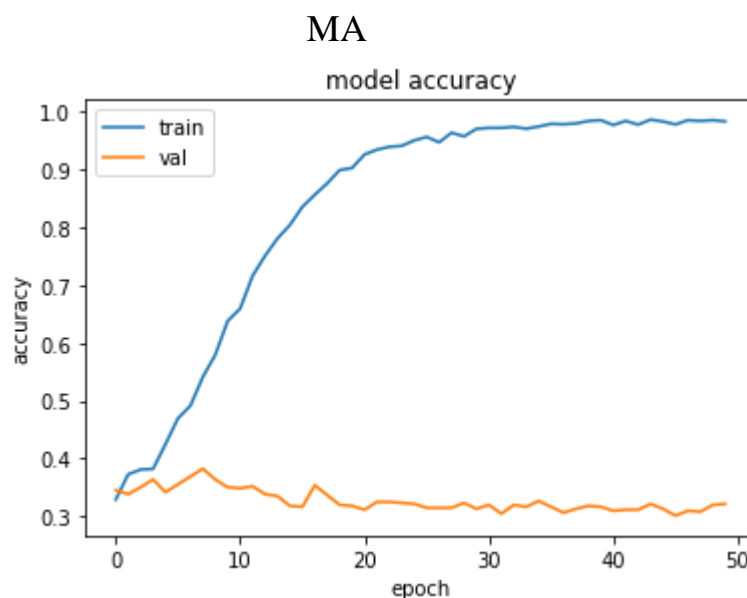


Рис. 10. График зависимости точности прогнозов МА для индекса S&P500 от числа эпох для обучающего и валидационного наборов.

Таблица А.2 – Значение метрик качества для метода МА и индекса S&P 500.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.33	0.41	0.37	200
1 (Промежуточное состояние)	0.43	0.43	0.43	200
2 (Рост)	0.41	0.31	0.35	200

Полученное значение для Accuracy: 0.38.

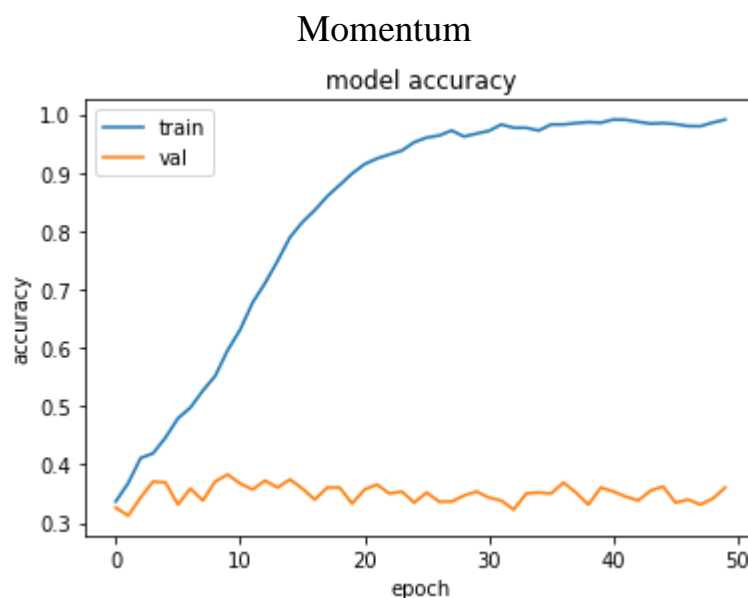


Рис. 11. График зависимости точности прогнозов Momentum для индекса S&P500 от числа эпох для обучающего и валидационного наборов.

Таблица А.3 – Значение метрик качества для метода Momentum и индекса S&P 500.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.41	0.32	0.36	200
1 (Промежуточное состояние)	0.40	0.41	0.41	200
2 (Рост)	0.35	0.41	0.38	200

Полученное значение для Accuracy: 0.38.

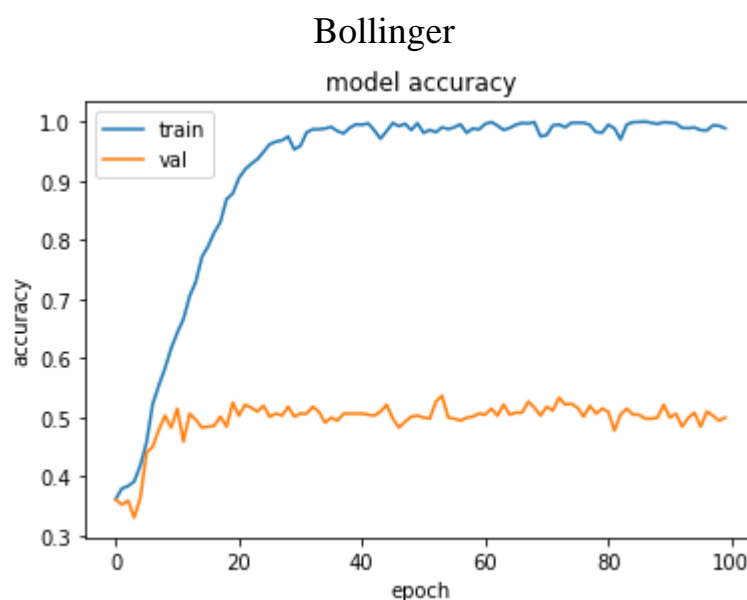


Рис. 12. График зависимости точности прогнозов Bollinger для индекса S&P500 от числа эпох для обучающего и валидационного наборов.

Таблица А.4 – Значение метрик качества для метода Bollinger и индекса S&P 500.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.60	0.71	0.65	200
1 (Промежуточное состояние)	0.34	0.41	0.37	200
2 (Рост)	0.62	0.35	0.45	200

Полученное значение для Accuracy: 0.49.

Приложение В. Результаты для индекса NASDAQ

GAF

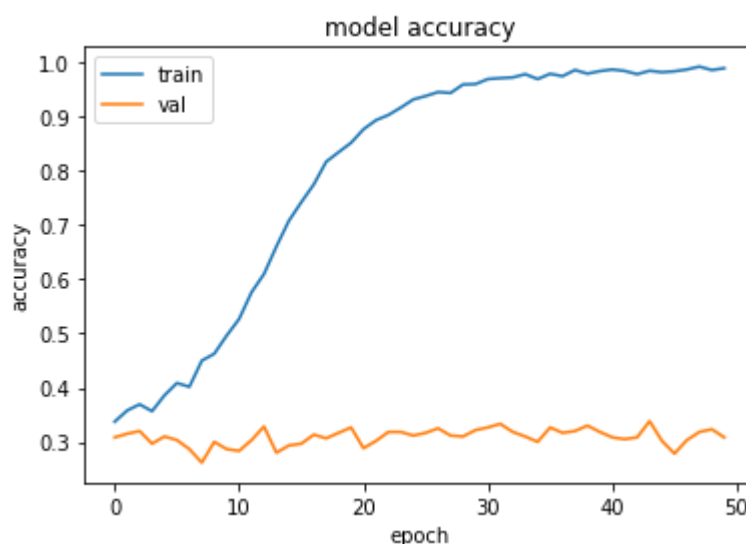


Рис. 13. График зависимости точности прогнозов GAF для индекса NASDAQ от числа эпох для обучающего и валидационного наборов.

Таблица В.1 – Значение метрик качества для метода GAF и индекса NASDAQ.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.38	0.40	0.39	200
1 (Промежуточное состояние)	0.40	0.35	0.37	200
2 (Рост)	0.46	0.50	0.48	200

Полученное значение для Accuracy: 0.42.

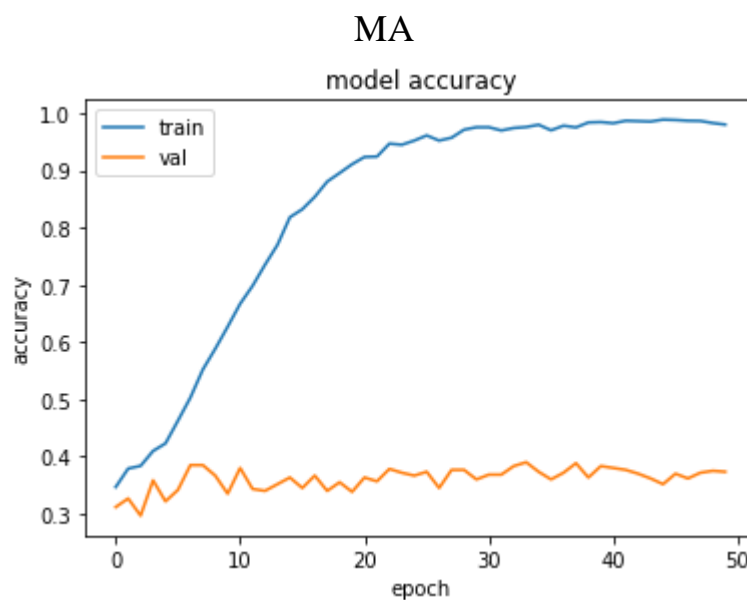


Рис. 14. График зависимости точности прогнозов МА для индекса NASDAQ от числа эпох для обучающего и валидационного наборов.

Таблица В.2 – Значение метрик качества для метода МА и индекса NASDAQ.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.32	0.32	0.32	200
1 (Промежуточное состояние)	0.32	0.31	0.32	200
2 (Рост)	0.34	0.36	0.35	200

Полученное значение для Accuracy: 0.33.

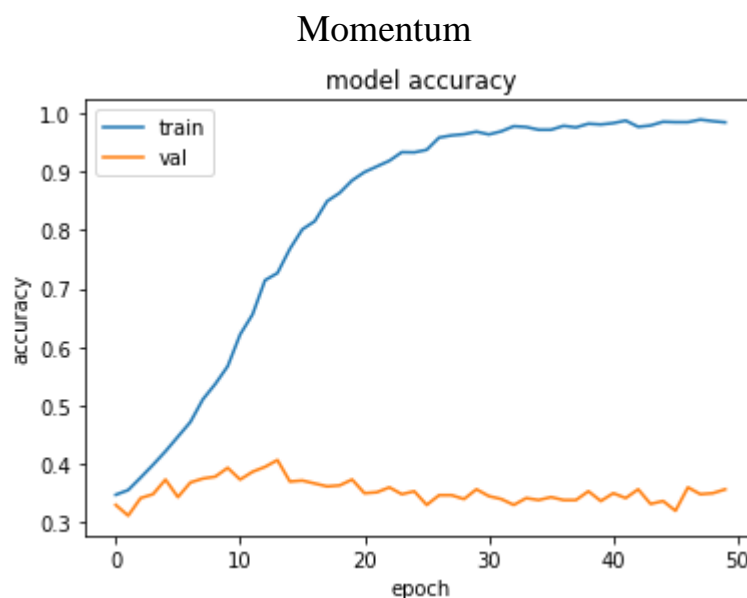


Рис. 15. График зависимости точности прогнозов Momentum для индекса NASDAQ от числа эпох для обучающего и валидационного наборов.

Таблица В.3 – Значение метрик качества для метода Momentum и индекса NASDAQ.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.32	0.35	0.34	200
1 (Промежуточное состояние)	0.41	0.28	0.34	200
2 (Рост)	0.38	0.47	0.42	200

Полученное значение для Accuracy: 0.37.

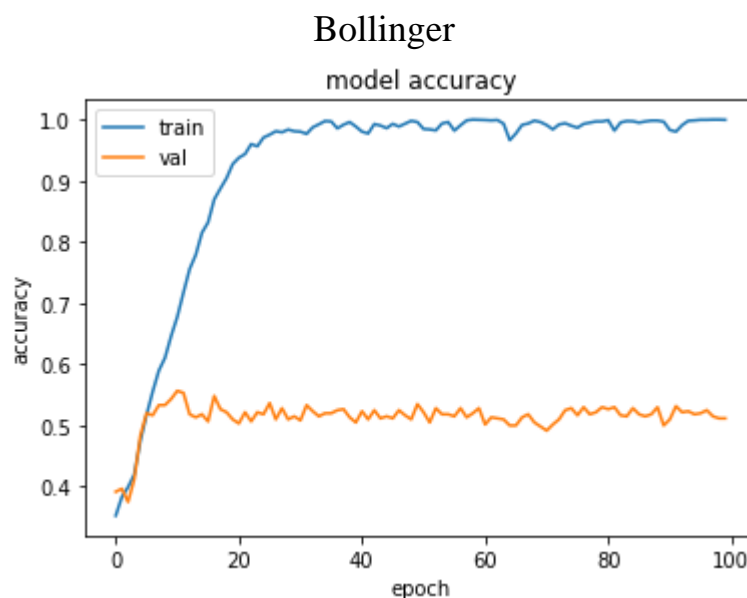


Рис. 16. График зависимости точности прогнозов Bollinger для индекса NASDAQ от числа эпох для обучающего и валидационного наборов.

Таблица В.4 – Значение метрик качества для метода Bollinger и индекса NASDAQ.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.62	0.86	0.72	200
1 (Промежуточное состояние)	0.37	0.28	0.32	200
2 (Рост)	0.57	0.48	0.52	200

Полученное значение для Accuracy: 0.54.

Приложение С. Результаты для индекса Dow Jones GAF

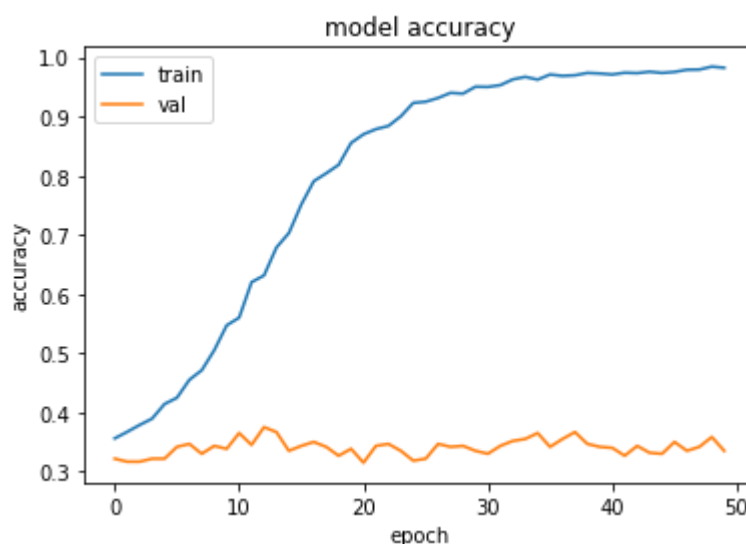


Рис. 17. График зависимости точности прогнозов GAF для индекса Dow Jones от числа эпох для обучающего и валидационного наборов.

Таблица В.5 – Значение метрик качества для метода GAF и индекса Dow Jones.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.31	0.29	0.30	200
1 (Промежуточное состояние)	0.33	0.43	0.37	200
2 (Рост)	0.36	0.27	0.31	200

Полученное значение для Accuracy: 0.33.

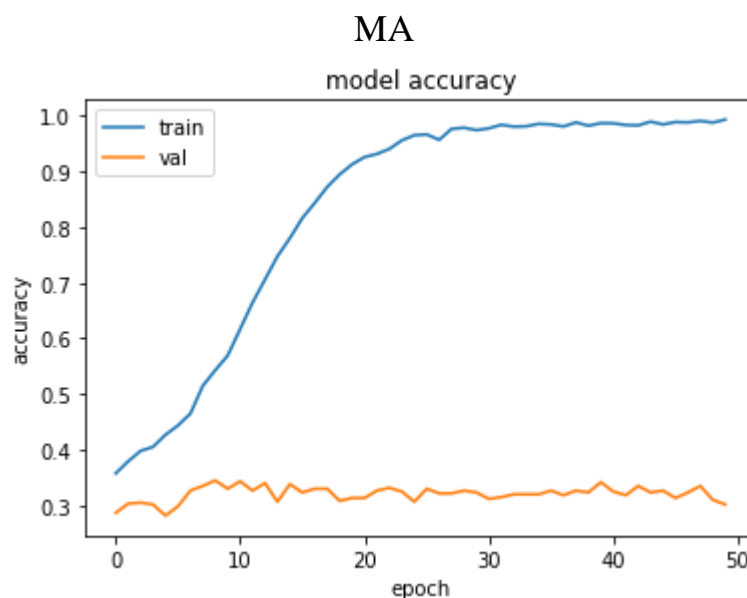


Рис. 18. График зависимости точности прогнозов МА для индекса Dow Jones от числа эпох для обучающего и валидационного наборов.

Таблица В.6 – Значение метрик качества для метода МА и индекса Dow Jones.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.34	0.52	0.41	200
1 (Промежуточное состояние)	0.43	0.40	0.41	200
2 (Рост)	0.38	0.20	0.27	200

Полученное значение для Accuracy: 0.37.

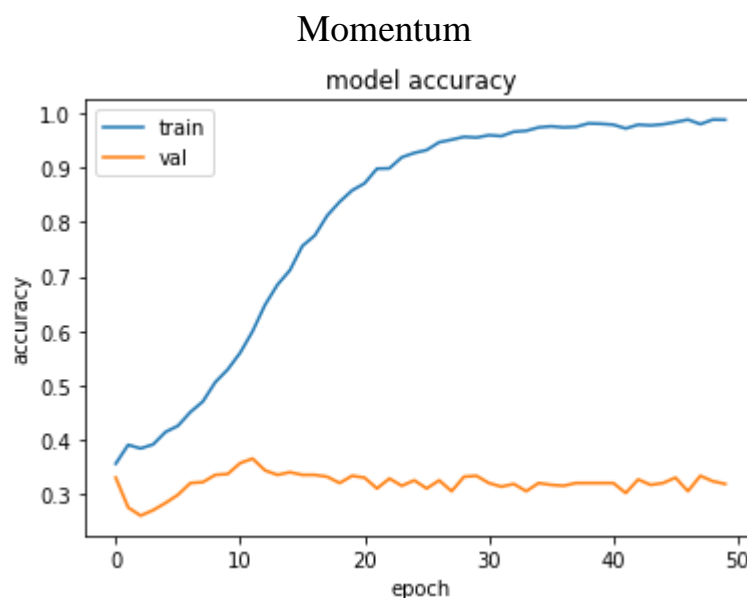


Рис. 19. График зависимости точности прогнозов Momentum для индекса Dow Jones от числа эпох для обучающего и валидационного наборов.

Таблица В.7 – Значение метрик качества для метода Momentum и индекса Dow Jones.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.39	0.40	0.39	200
1 (Промежуточное состояние)	0.44	0.60	0.51	200
2 (Рост)	0.42	0.26	0.32	200

Полученное значение для Accuracy: 0.42.

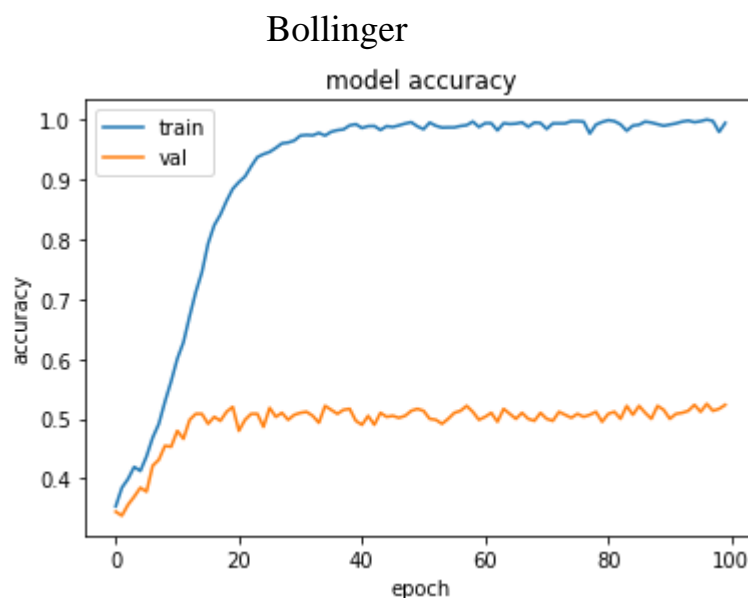


Рис. 20. График зависимости точности прогнозов Bollinger для индекса Dow Jones от числа эпох для обучающего и валидационного наборов.

Таблица В.8 – Значение метрик качества для метода Bollinger и индекса Dow Jones.

Класс	Precision	Recall	F1-score	Количество примеров
0 (Падение)	0.52	0.74	0.61	200
1 (Промежуточное состояние)	0.30	0.32	0.31	200
2 (Рост)	0.47	0.25	0.33	200

Полученное значение для Accuracy: 0.43.